Newsroom OnTAP – Real-time alerting from streaming audio

Scott Shepard, Sean Colbath, Dr. Kathleen Egan[1], *Francis Kubala*

BBN Technologies, Cambridge, MA 02138
sshepard@bbn.com

[1] Department of Defense

1. INTRODUCTION

The Newsroom OnTAP system is an operational system that permits automatic monitoring of several continuous data streams for detection of speech passages that match user-defined profiles. The Newsroom OnTAP system is an example of an OnTAP system which means that it employs a diverse set of human language technologies for processing a large and continuous stream of speech and text data from diverse media sources.

BBN's OnTAP system, which stands for *Online Text and Audio Processing*, is an experimental system under development in the DARPA TIDES program. The Newsroom OnTAP system is to be fielded in stages beginning in February 2002. In this demo, we show the capabilities of the Newsroom OnTAP system that provide real-time alerting of user-defined subjects and events from continuous data streams.

2. REAL-TIME ALERTING

Real-time alerting is the process of notifying a user that data related to specific profiles, or user-defined criteria, has been detected. The time between the presence of the data and the user notification is critical and should be minimized. There are two forms of profile specifications: keyword and event. A keyword profile enables the user to query the continuous data streams for specific combinations of keywords, similar to a typical search engine query but a keyword profile persists and is applied continuously to incoming data. Event based profiles enable the user to define a target event by one or more example documents. By combining these two forms of real-time alerting, the system provides the user with a flexible way of defining data of interest while providing real-time notification when relevant data has arrived.

3. NEWSROOM ONTAP

The Newsroom OnTAP system provides the user with a means of monitoring more data streams than previously possible. By automatically monitoring multiple data streams in the background and only notifying the user when data of particular interest is detected, the user can focus his attention on only data that is of particular interest to him.

The Newsroom OnTAP system architecture consists of one or more BBN Oasis Audio Indexers, one or more Event Trackers, and the Newsroom Server, as shown in Figure 1. The user's desktop is only required to have a supported browser (IE 5.0+) and Real Networks RealPlayer installed. The Newsroom OnTAP system handles speech and text interchangeably.

The Oasis Audio Indexers produce real-time transcriptions of data from the continuous data streams. They publish these transcriptions to the Event Trackers and the Alerting Service. The Event Trackers perform an incremental tracking operation on the data provided by the Oasis Audio Indexers and report the results to the Alerting Service. The Alerting Service compares the data provided by the Oasis Audio Indexers and Event Trackers to the user-defined alerting criteria. If the data satisfies the alerting criteria, the data is stored in the alerting database and the user is prompted with an alert via the web. The alert will appear in the users web browser without any user interactions, it is not necessary for the user to refresh the web page.

All user interactions with the Newsroom OnTAP system are accomplished via a Web browser. The user can create new alert criteria as well as receive alert notifications via the web. When the user receives an alert notification, they have the ability to view the relevant data immediately. If the data stream

was an audio broadcast, the user will see a transcription of the broadcast with the relevant words highlighted. The user has the ability to replay the audio for the relevant portion of the broadcast directly from their web browser. If the data stream is text, the user will see the complete document with all of the relevant words highlighted in the web browser.

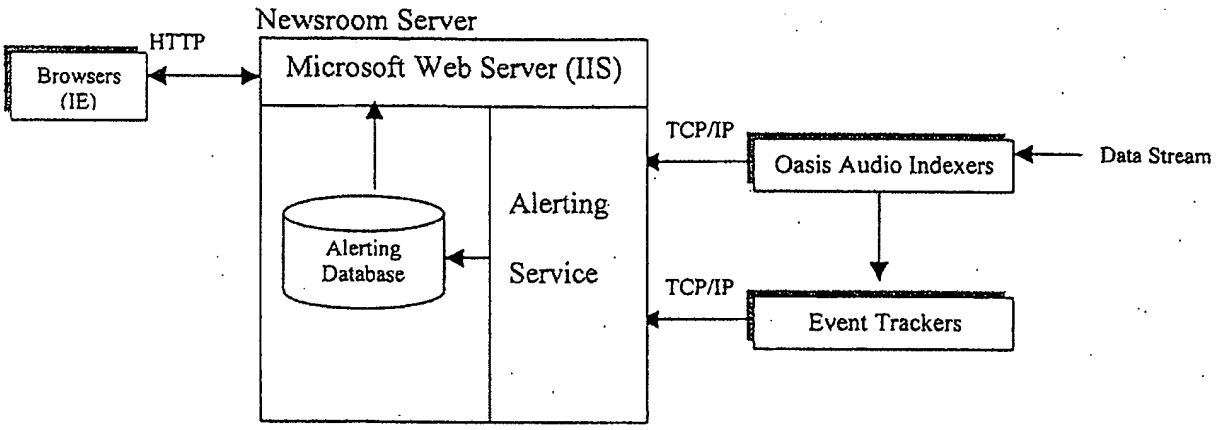

Figure 1: Newsroom OnTAP System diagram

4. DEMONSTRATION

The HLT demonstration system will monitor one continuous data stream throughout the demonstration period for real-time alerting. The audience will be encouraged to set up profiles throughout the conference and experience the technologies offered by Newsroom OnTap.